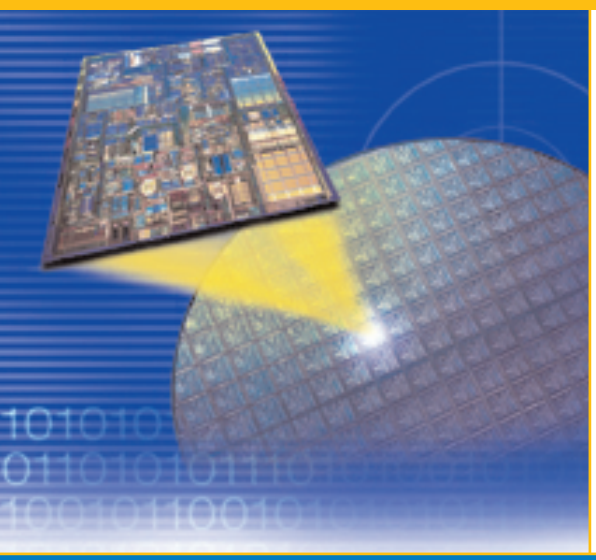
Volume 06 Issue 02 Published, May 16, 2002 ISSN 1535766X



Intel® Technology Journal

Semiconductor Technology and Manufacturing

130nm Logic Technology Featuring 60nm Transistors, Low-K Dielectrics and Cu Interconnects

A compiled version of all papers from this issue of the Intel Technology Journal can be found at:
http://developer.intel.com/technology/itj/index.htm

130nm Logic Technology Featuring 60nm Transistors, Low-K Dielectrics, and Cu Interconnects

Scott Thompson, Technology and Manufacturing Group, Intel Corporation
Mohsen Alavi, Technology and Manufacturing Group, Intel Corporation
Makarem Hussein, Technology and Manufacturing Group, Intel Corporation
Pauline Jacob, Technology and Manufacturing Group, Intel Corporation
Chris Kenyon, Technology and Manufacturing Group, Intel Corporation
Peter Moon, Technology and Manufacturing Group, Intel Corporation
Matthew Prince, Technology and Manufacturing Group, Intel Corporation
Sam Sivakumar, Technology and Manufacturing Group, Intel Corporation
Sunit Tyagi, Technology and Manufacturing Group, Intel Corporation
Mark Bohr, Technology and Manufacturing Group, Intel Corporation

Index words: CMOS transistor, logic technology, copper interconnects

ABSTRACT

Transistor gate dimensions have been reduced 200X during the past 30 years (from 10 m in the 1970s to a present-day size of 0.06 m). The transistor and feature size scaling have enabled microprocessor performance to increase exponentially with transistor density and microprocessor clock frequency doubling every two years. In this paper we describe Intel's latest 130nm CMOS logic technology used to make high-performance microprocessors >3GHz.

INTRODUCTION

For more than 30 years, MOS device technologies have been improving at a dramatic rate [1-6]. A large part of the success of the MOS transistor is due to the fact that it can be scaled to increasingly smaller dimensions, which results in higher performance. The ability to consistently improve performance while decreasing power consumption has made CMOS architecture the dominant technology for integrated circuits. The scaling of the CMOS transistor has been the primary factor driving improvements in microprocessor performance. Transistor delay times have decreased by more than 30% per technology generation resulting in a doubling of microprocessor performance every two years. Recently, chip performance has also come to be limited by back-end RC delay if low-resistance metal lines or low dielectric constant interlayer dielectrics are not used.

In this paper we describe Intel's 130nm logic technology that features 60nm gate length and 1.5nm gate-oxide transistors for high-performance and low-k interdielectrics with six layers of Cu interconnects. We first discuss transistor scaling. Next, we present data from our 130nm technology on 60nm transistors and copper interconnects with low-k Fluorinated SiO_2. We conclude with static random access memory (SRAM) and microprocessor performance data.

TRANSISTOR SCALING OVERVIEW

Transistor scaling has been the key driving force behind the rapid increase in microprocessor clock frequency. Figure 1 shows the scaling trend of clock frequency. The technology target for the 130nm node was to produce microprocessors at >3GHz.

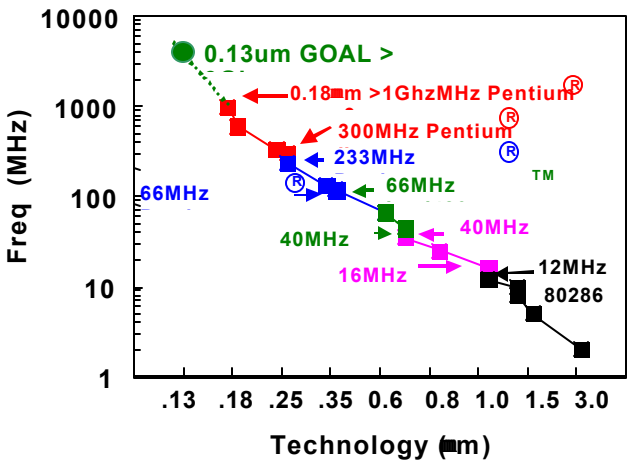


Figure 1: Microprocessor clock frequency vs. technology generation

Two primary factors set the transistor speed and hence microprocessor clock frequency for a given design: transistor channel length and gate-oxide thickness (Figure 2). To reach the >3GHz goal, circuit simulations show that 60nm gate length and 1.5nm gate-oxide thickness are required for the 130nm technology node. The 60nm transistor requires a significant acceleration of the transistor feature size relative to the technology and light source.

Figure 3 shows the trends of these key feature sizes versus technology generation. The 130nm technology node was designed for the fabrication of Intel Pentium® 4 microprocessors in high-volume manufacturing. Once the Pentium 4 chip architecture is set, the transistor speed required for 3GHz operation can be determined. To obtain a clock frequency of >3GHz, it was determined that a 1.3mA/ m transistor saturation drive current would be needed. This value of drive current is significantly higher than the value in our 180nm technology (~1.0mA / m). 60nm transistors with 1.5nm physical oxide thickness will allow for CV/I close to 1ps (Figure 4) and saturation drive current of 1.3mA/ m (Figure 6). This drive current is the highest to date in high-volume production. Key to obtaining the high drive current is high channel mobility. The channel mobility decreases at higher effective oxide fields for the smaller feature size technology. The electron mobility is shown in Figure 5. The electron mobility is on the universal mobility curve even though the physical thickness of the oxide is only 1.5nm.

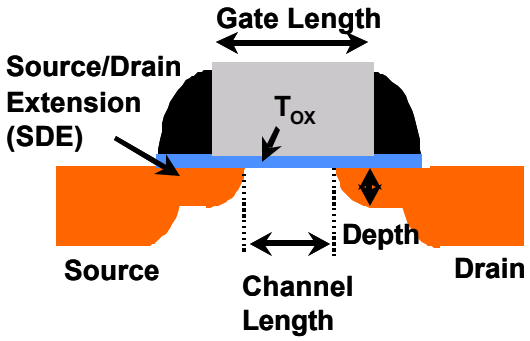


Figure 2: Cross-section drawing of a CMOS transistor

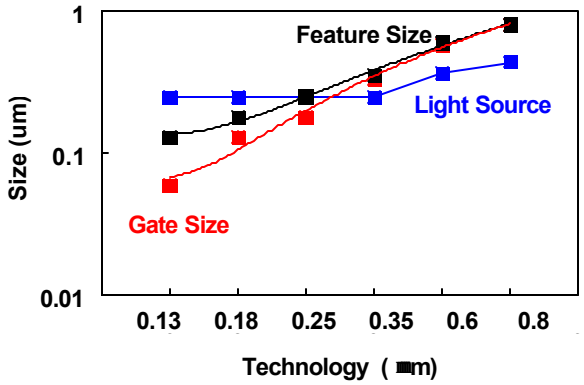


Figure 3: Technology feature size, wavelength light source, and transistor gate size vs. technology node

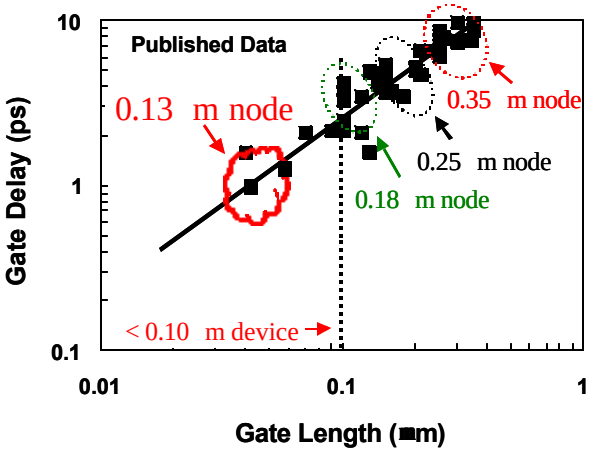


Figure 4: CV/I gate delay vs. transistor gate length

Intel and Pentium are registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries.

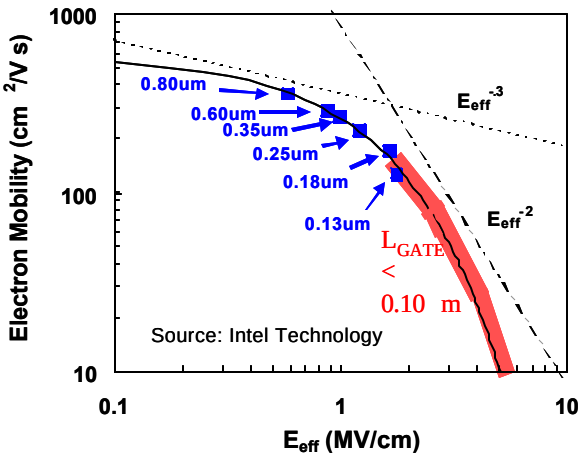


Figure 5: Electron mobility vs. effective vertical electrical field

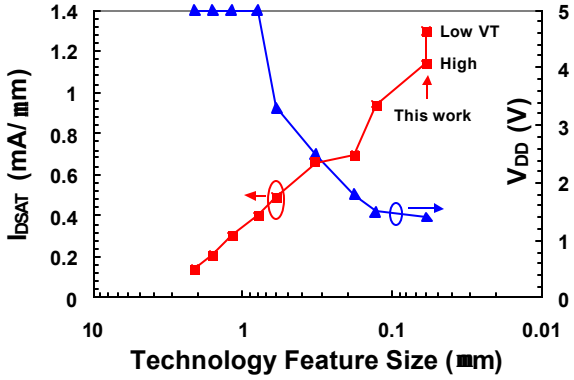


Figure 6: Transistor saturated drive current vs. technology feature size

Process Flow and Technology Features

Front-end technology features include shallow trench isolation, retrograde wells, shallow abrupt source/drain extensions, halo implants, deep source/drain, and cobalt salicidation. Figure 7 shows a front-end cross section of the technology. The minimum pitches and thicknesses for the technology layers are summarized in Table 1. The rules enable a 2.0 um ^{2} 6-T SRAM cell (1.22 x 1.64 um). Figure 8 shows a top-down scanning electron micrograph (SEM) of the polysilicon gate conductor and the Metal 1 connections. The interconnect technology uses dual damascene copper to reduce the resistances of the six layers of interconnects. Fluorinated SiO_2 is used as an inter-level dielectric (k is measured to be 3.6).

Table 1: Layer pitch, thickness (nm) and aspect ratio

LAYER	PITCH	THICK	AR
Isolation	345	450	-
Polysilicon	319	160	-
Metal 1	293	280	1.7
Metal 2,3	425	360	1.7
Metal 4	718	570	1.6
Metal 5	1064	900	1.7
Metal 6	1143	1200	2.1

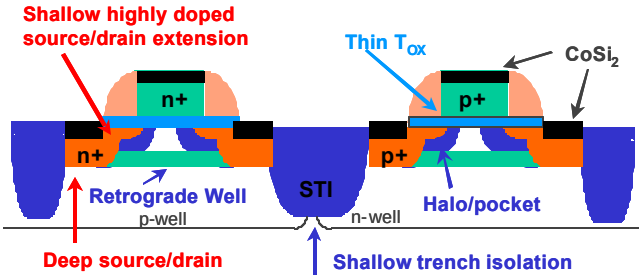


Figure 7: Cross-section drawing of 130nm technology front-end

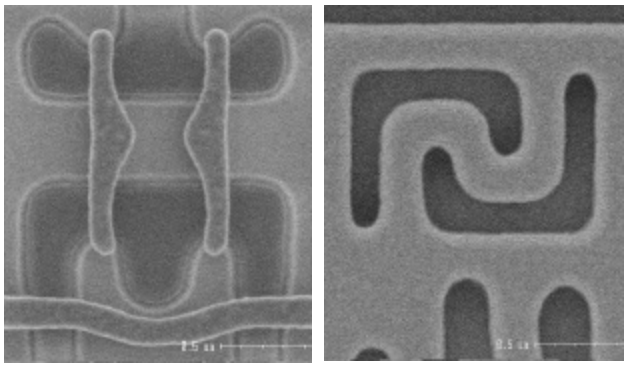


Figure 8: Top-down SEM of polysilicon gate conductor and Metal 1 connections

TRANSISTOR FEATURES

(a) Gate Length Dimension: Figure 9 shows a cross-sectional transmission electron micrograph (TEM) for a transistor with a 60nm gate length and straight poly-Si sidewall profile as opposed to the notched poly used in the 180nm node [6]. Straight sidewall gates were chosen at the 130nm node since the source drain extension does not have to diffuse under the notch, thus allowing for shallower junctions to be fabricated.

At aggressive gate lengths of 60nm, controlling short channel effects at low-threshold voltage, by using shallow junctions and abrupt halo doping, is key to achieving high linear and saturation drive currents.

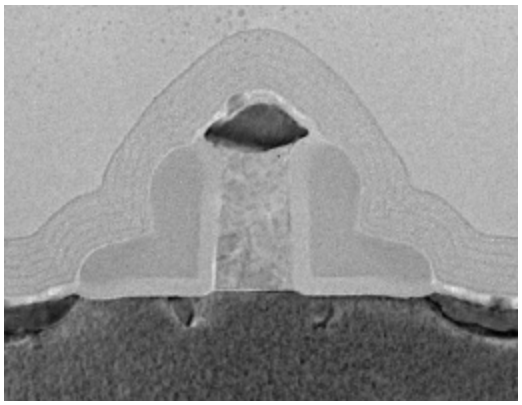


Figure 9: TEM cross section of 60nm NMOS

(b) 1.5nm Physical Gate-Oxide: In order to achieve high drive current and minimize short channel effects, a gate-oxide process with a 1.5nm physical thickness was developed that meets performance, reliability, and manufacturability criteria (Figure 10). High-electron and hole mobilities are required to achieve high linear drive current, which can be missed in technology optimization, since transistor linear current is not reflected in a simple CV/I metric. Concerns have been raised that in ultra-thin oxides, gate-electrode-to-oxide interface scattering and high fixed charge due to nitridation reduce mobility. The measured mobility dependence on the effective oxide field, shown in Figure 11, demonstrates that high-electron and hole mobilities can be achieved for well-optimized gate oxides with a thickness of 1.5nm.

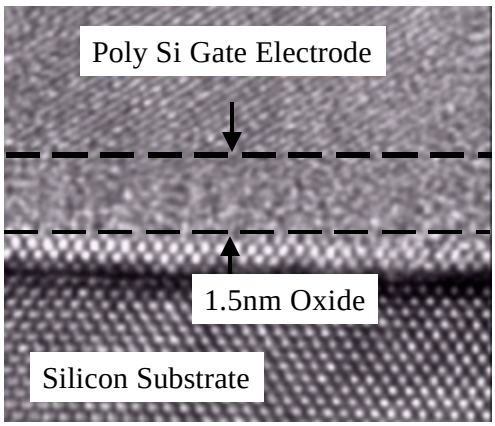


Figure 10: TEM of 1.5nm physical gate oxide

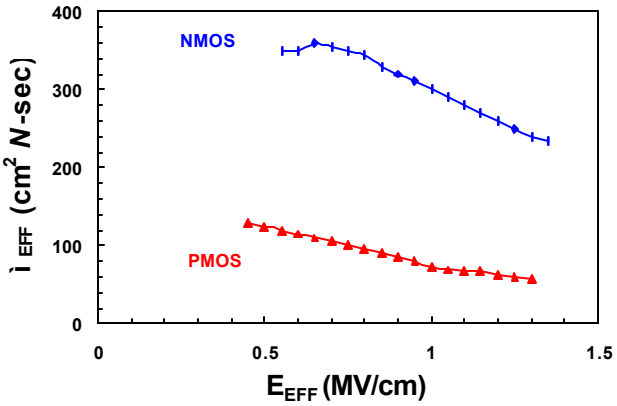


Figure 11: Mobility vs. effective electric field

(c) Well-Halo and SD-Extension Engineering: A simple but ineffective way to offer high-saturation drive current at small gate lengths is to use high well-doping to raise the threshold voltage to control short channel effects. This approach offers low CV/I but does not improve product performance, for two reasons. First, the linear drive current will be significantly degraded (saturated drive current is not degraded at a fixed I_{OFF} due to high drain-induced barrier lowering (DIBL)). Second, the high well-doping leads to increased threshold voltage variations due to gate length variation (present in the range of +/- 10% L_{GATE} for a modern technology). In this work we use retrograde wells, and low-energy, high-angle abrupt halo implants with shallow junctions formed by low-diffusion processing to control short channel effects. Figure 12 shows the N-channel threshold voltage versus gate length resulting in a linear threshold voltage of 300 and 270mV at a gate length of 60nm for the high- and low-threshold devices, respectively. From Figure 12, DIBL for the 60nm NMOS devices is measured to be <100mV/V for high- and bw-threshold devices. Similar results have been achieved for p-channel devices.

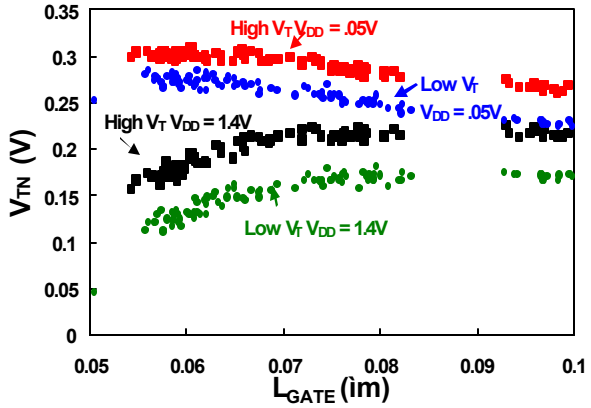


Figure 12: V_{TN} vs. L_{GATE}

High V_T saturation drive currents are 1.14mA/um for N-channel and 0.56mA/um for P-channel devices (Figure 13). Low V_T drive currents are 1.30mA/um for N-channel and 0.66mA/um for P-channel devices (Figure 14). Sub-threshold slopes for both N-channel and P-channel high- and low-threshold devices remain well controlled at less than 85mV/decade at L_{GATE}=60 nm (Figure 15). The I_{ON}/I_{OFF} ratio remains high for the aggressively scaled power supply voltage of 1.4V (Figure 16). Table 2 shows the transistor I_{ON} and I_{OFF} at 0.7 and 1.4 V.

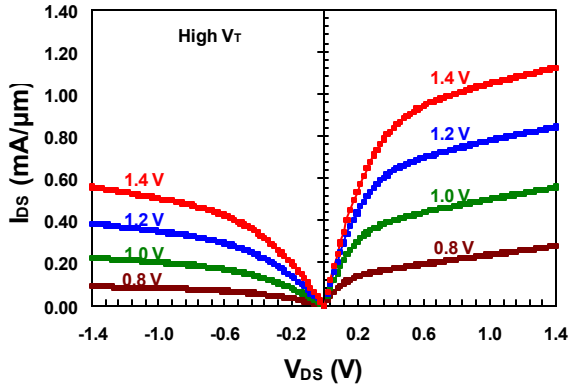


Figure 13: I-V curves for high V_T device (L_{GATE}=60nm)

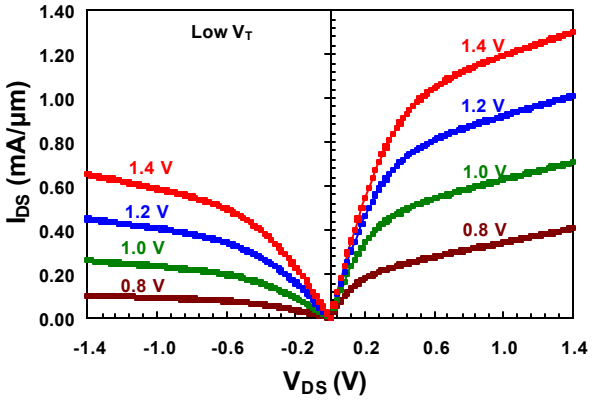


Figure 14: I-V curves for low V_T device (L_{GATE}=60nm)

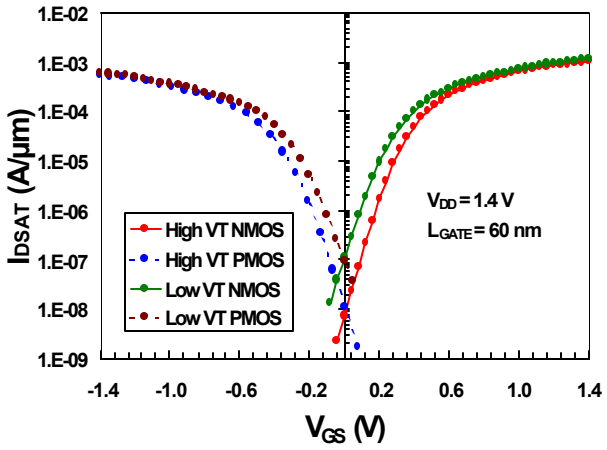


Figure 15: Sub-threshold characteristics

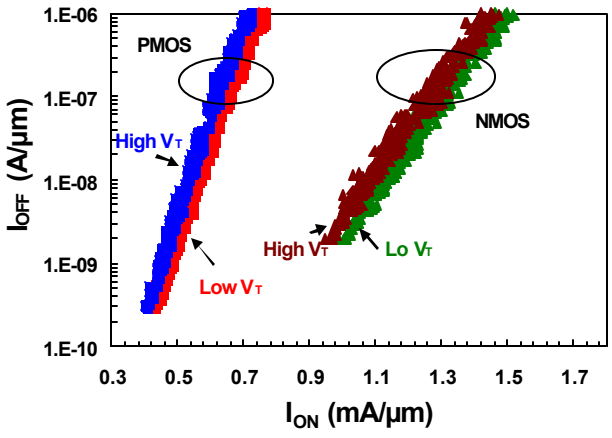


Figure 16: I_{ON} Vs I_{OFF} (V_{DD}=1.4V)

Table 2: I_{ON} and I_{OFF} at 0.7 and 1.4V V_{DD}

DEVICE	VDD (V)	I_{OFF} (N) (nA/um)	I_{ON} (N) (mA/um)	I_{ON}(P) (mA/um)
Low V_T	1.4	100	1.30	0.66
High V_T	1.4	10	1.14	0.56
Low V_T	0.7	20	0.37	0.19
High V_T	0.7	2	0.32	0.16

In a modern microprocessor with six layers of interconnects, transistor loads are comprised of >50% interconnect capacitance. To obtain high product performance it is necessary to provide transistors with more than low CV/I; you also need high saturation and linear drive currents. Figure 6 shows the recent trend of saturation drive currents for Intel's process technologies. This work extends the trend to offer the highest drive current to date of 1.30mA/um for low-threshold N-channel devices.

INTERCONNECTS

Chip performance is increasingly limited by the RC delay of the interconnect as the transistor delay progressively decreases while the narrower lines and space actually increase the delay associated with interconnects. Using copper interconnects helps reduce this effect. This process technology uses dual damascene copper to reduce the resistances of the interconnects. Fluorinated SiO_2 (FSG) is used as an inter-level dielectric (ILD) to reduce the dielectric constant; the dielectric constant k is measured to be 3.6. Figure 17 is a cross-section Scanning Electron Micrograph (SEM) image showing the dual damascene interconnects.

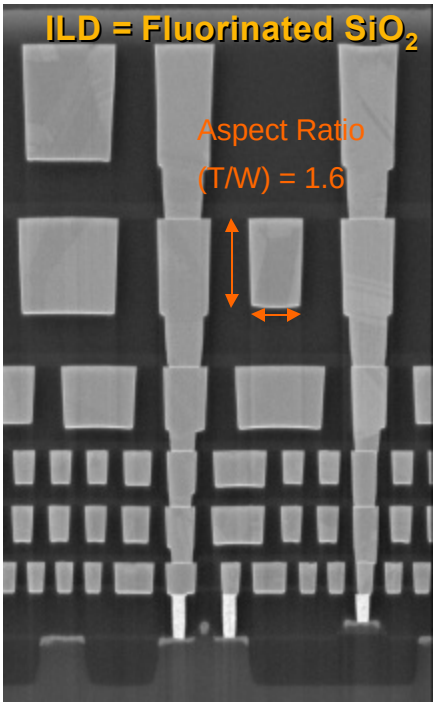


Figure 17: Cross-section SEM image of a processed wafer

Table 1 lists the metal pitches. The pitch is 350nm at the first metal layer and increases to 1200nm at the top layer. Metal aspect ratios are optimized for minimum RC delay and range from 1.6 to 2. The first metal layer uses a single damascene process, and tungsten plugs are used as contacts to the silicided regions on the silicon and poly-silicon. Unlanded contacts are supported by using an Si_3N_4 layer for a contact etch stop. Copper interconnects are used because of the material's lower resistivity. The advantage is seen in Figure 18, where the sheet resistance is shown as a function of the minimum pitch of each metal layer and compared to earlier results from 180nm technologies using Al [6] and Cu [6]. The present technology exhibits 30% lower sheet resistance at the same metal pitch due to the use of Cu with high aspect ratios. The total line capacitance is 230fF/mm for M1 to M5 and slightly higher for the top layer.

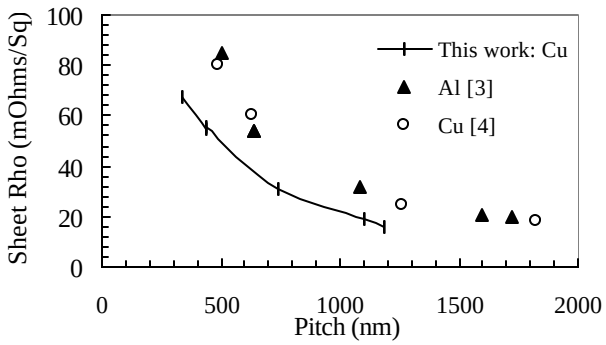


Figure 18: Sheet resistance as a function of layer pitch

To benchmark the performance of interconnects, Figure 19 shows the RC delay in picoseconds per millimeter of wire. Data for each metal layer are shown as a function of the minimum pitch at that layer. For a given pitch, 50% reduction in RC is achieved by using Cu interconnects and FSG ILD.

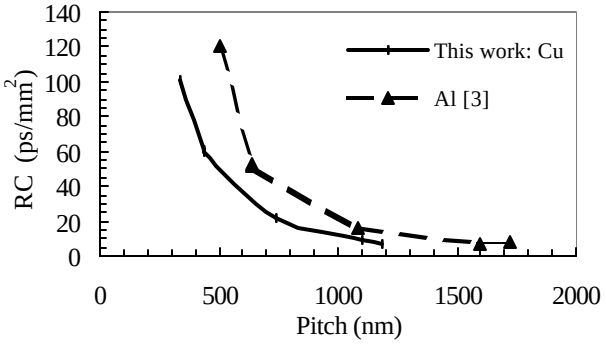


Figure 19: RC delay for a wire length of 1mm as a function of layer pitch

Performance Metrics

Figure 20 shows measured inverter gate delay versus n-channel off-state leakage for an unloaded ring oscillator (fan out =1) operating at 1.4V at room temperature. PMOS off-state leakage is fixed at 10nA/um for these devices. The delay per stage at 1.4V falls below 6psec when the off-state leakage is about 10nA/um.

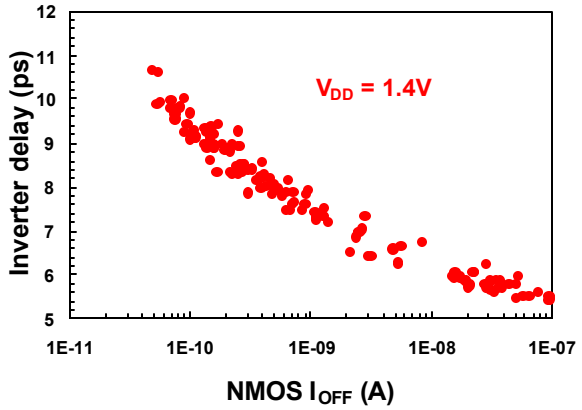


Figure 20: Inverter delay (PMOS I_{OFF} = 10nA/um)

Power consumption is a growing concern for high-performance microprocessors with increasing clock frequency and transistor count. The best way to reduce power is to operate at a low supply voltage. Figure 21 shows that by improving device matching and eliminating defects that cause device mismatches, an 18Mb SRAM fabricated in this technology can operate at voltages of down to 0.5V. A metric, which comprehends both power and speed, is the energy-delay product. Figure 22 shows the estimated NMOS energy-delay product for a large number of published devices and for the devices reported in this paper. As evident from Figure 22, the NMOS energy-delay product is better than the published industry trend.

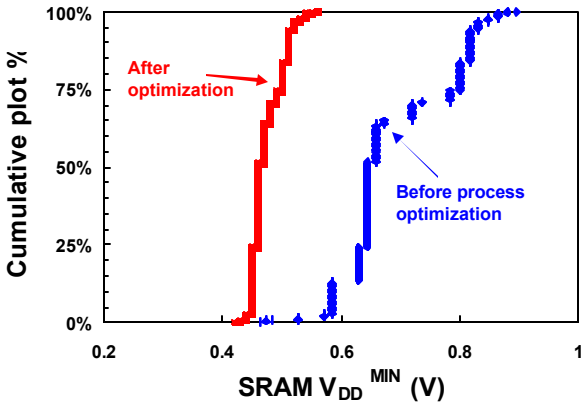


Figure 21: SRAM operation vs. voltage

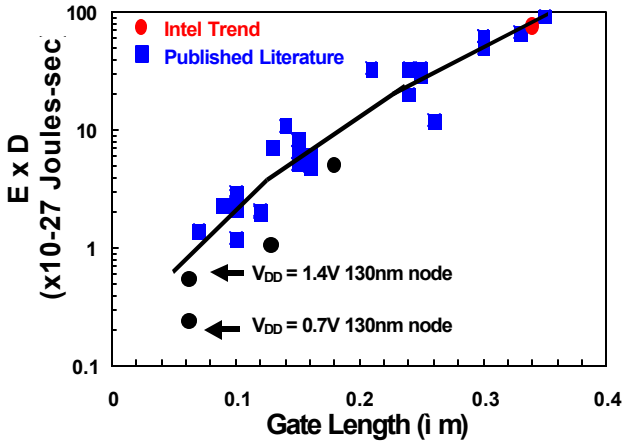


Figure 22: Energy-delay product vs. L_{GATE}

An 18 Mbit CMOS SRAM, Pentium III and Pentium® 4 microprocessor were fabricated and used as yield and reliability test vehicles during the process development. Figure 23 shows the die photo of the Pentium 4 in the 0.18 and 0.13um technologies. The SRAM and microprocessor die yields are equivalent or better than past technologies at this point of time relative to ramping in high-volume manufacturing. The performance of the Pentium 4 processor is measured using the maximum clock frequency of operation. Figure 24 shows the schmoo plot for the Pentium 4, i.e., the maximum frequency as a function of voltage. At an operation voltage of 1.4V, the present design version of the Pentium 4 microprocessor has a clock frequency of 2.5GHz.

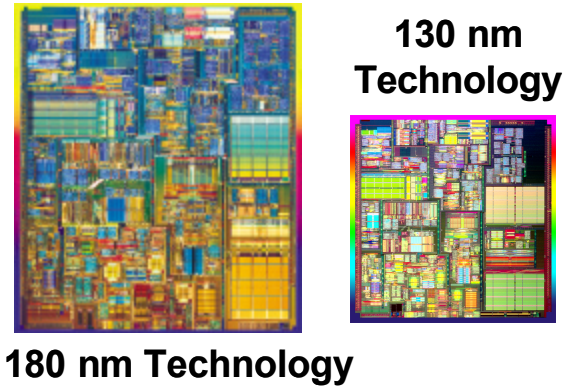


Figure 23: Comparison of 180nm technology to 130nm technology

Pentium is a registered trademark of Intel Corporation or its subsidiaries in the United States and other countries.

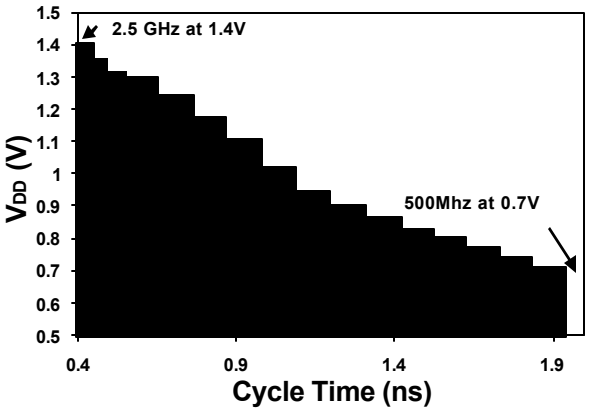


Figure 24: Fmax schmoo plot for the Pentium® 4 processor

CONCLUSIONS

A 130nm-generation logic technology has been developed and is in high-volume manufacturing with high-performance transistors that can operate in the range of 0.7 and 1.4 V. The technology performance capabilities are demonstrated with ring oscillator delays of 6 ps/stage and with a Pentium 4 processor operating at 2.5 GHz. The transistors can support microprocessors operating at >3GHz.

ACKNOWLEDGMENTS

The authors acknowledge the collaborative efforts of our colleagues in the Portland Technology Development Group, the Technology Computer Aided Design Group, and in the Corporate Quality and Reliability group.

REFERENCES

[1] M. Bohr, S.U. Ahmed, L. Brigham, R. Chau, R. Gasser, R. Green, W. Hargrove, E. Lee, R. Natter, S. Thompson, K. Weldon and S. Yang, *IEDM Technical Digest*, 1994, p. 273.

[2] M. Bohr, S.S. Ahmed, S.U. Ahmed, M. Bost, T. Ghani, J. Greason, R. Hainsey, C. Jan, P. Packan, S. Sivakumar, S. Thompson, J. Tsai, and S. Yang, *IEDM Technical Digest*, 1996, p. 847.

[3] S. Thompson, VLSI Symposium Technology Short Course, 1998.

[4] S. Thompson, P.A. Packan, and M.T. Bohr, *VLSI Symposium Digest*, 1996, p. 154.

[5] C.T. Sah, Fundamentals of Solid-State Electronics, 1991, p. 553.

[6] S. Thompson, M. Alavi, R. Arghavani, A. Brand, R. Bigwood, J. Brandenburg, B. Crew, V. Dubin, M. Hussein, P. Jacob, C. Kenyon, E. Lee, B. Mcintyre, Z. Ma, P. Moon, P. Nguyen, M. Prince, R. Schweinfurth, S. Sivakumar, P. Smith, M. Stettler, S.Tyagi, M. Wei, J. Xu, S. Yang and M. Bohr, *IEDM Technical Digest*, 2001, p. 11.6.1-11.6.4.

AUTHORS' BIOGRAPHIES

Scott Thompson joined Intel in 1992 after completing his Ph.D., under Professor C. T. Sah at the University of Florida, on thin gate oxides. He has worked on transistor design and front-end process integration on Intel's 0.35, 0.25, 0.18, and 0.13 m silicon process technology design for the Intel® Pentium® and the Pentium® II microprocessors. Scott is currently managing the development of Intel's 90nm logic technology. His e-mail is scott.thompson@intel.com.

Mohsen Alavi joined Intel in 1986 after completing his Ph.D. in Electrical Engineering at Michigan State University on Schottky Barrier Diodes. He has worked on transistor and interconnect development and reliability of many of Intel's logic process technologies starting from the 1um process. More recently, he has been the reliability program manager for 0.13um and subsequently, 90nm logic technology development and is currently manager of LTD Q&R. His e-mail is mohsen.alavi@intel.com.

Makarem Hussein is a Principal Engineer with Patterning Area of Portland Technology Development. He graduated from the University of Wisconsin-Madison in 1990 with a Ph.D. degree in Nuclear Engineering and Engineering Physics. He joined Intel in 1992, and since then has been working on developing dry etch processes. His most recent focus has been on the patterning of dielectric substrates for copper interconnect systems. He holds six US patents and has authored/co-authored more than 15 articles in the field of plasma etching and patterning technology. His e-mail is makarem.hussein@intel.com

Pauline Jacob joined Intel in 1994 after completing a Ph.D. in Chemical Engineering at the University of Washington. She has worked on diffusion process development since Intel's 0.35um process technology. Pauline is currently the diffusion group leader working on the development of Intel's 90nm gate-oxide module. Her e-mail is pauline.n.jacob@intel.com.

Chris Kenyon is a lithography group leader in Intel's Logic Technology Development organization. He joined Intel in 1996 and has worked primarily on Intel's gate patterning process since that time. He is currently responsible for developing the gate patterning process for the 90nm CMOS node. He received his B.A. degree from Princeton University in 1990 and his Ph.D. degree from Caltech in 1996 in Physical Chemistry. His e-mail is Chris.Kenyon@intel.com

Peter Moon joined Intel in 1988 after completing his Ph.D. in Materials Science at the Massachusetts Institute of Technology. He has worked on process integration for Intel's 0.8, 0.35 and 0.13um silicon process technologies for Pentium® microprocessors including Intel's first use of shallow trench isolation (0.35um) and Intel's first use of copper interconnects (0.13um). Peter is currently leading the development of Intel's interconnect process for the 45nm process generation. His e-mail is Peter.Moon@intel.com

Sam Sivakumar joined Intel in 1990 after graduating from the University of Illinois. He is a member of the Portland Technology Development lithography group and has worked on patterning process development for a variety of Intel's logic processes. He is currently responsible for lithography development for Intel's 90nm logic process. His e-mail is sam.sivakumar@intel.com.

Mark T. Bohr joined Intel in 1978 after receiving an M.S.E.E. degree from the University of Illinois. He has been a member of the Portland Technology Development group since 1978 and has been responsible for process integration and device design on a variety of DRAM, SRAM, and logic technologies, including recent 0.35um and 0.25um logic technologies. He is an Intel Fellow and director of process architecture and integration. He is currently directing development activities on 0.18um and 0.13um logic technologies. His e-mail is mark.bohr@intel.com.

Matthew Prince joined Intel in 1988 after graduating from Clarkson University in New York. Since 1989 he has developed ILD, W, and Cu CMP technologies. His e-mail address is matthew.j.prince@intel.com.

For further information visit:

developer.intel.com/technology/itj/index.htm